Buros Center for Testing - CCC Advisory Assessment Committee Training (September 2017)
Sample examples created for training purposes based on fictitious data

CCC Standards: Reliability

- Sample representative of CCC students (Minimum 50)
- Report reliability estimate and standard error measurement(SEM)
- Reliability estimates minimums
    - Test-retest (1 form, 2 occasions) = .75 or higher
    - Internal consistency (1 form, 1 occasion) = .80 or higher
    - Equivalent form/inter-prompt (2 forms, 1 occasion) = .75 or higher
    - Percent agreement (Human Scorers) = 90% or higher (1 point difference)
        - Report how to resolve inconsistencies between scorers

Other review considerations

a. Did reliability studies include a large enough sample of students that are representative of the CCC student population?
b. Did the lapse in time between test-retest testing occasions seem appropriate?
c. Are the standards errors of measurement (SEM) reasonable across the score scale, especially at likely cut points?

Test-Retest Reliability

<table>
<tr><td>

<u>Description of the Example</u>

In one CCC, ESL students who were beginning first-time enrollment completed the test once on the first day of a five-day orientation for ESL students. In the intervening days of the orientation, students met with teachers and with other students, took tours of campus, and met with their assigned advisor. No explicit English language instruction occurred. On the fifth day of the orientation, students again took the test. There is only one form of the test. A total of 68 students completed both tests; all were ESL students. The correlation between each student's scale score on the two tests was calculated as the measure of test-retest reliability. The calculated test-retest reliability (i.e., the correlation between the two administrations) was .77.

</td></tr>
</table>

Buros Center for Testing - CCC Advisory Assessment Committee Training (September 2017)
Sample examples created for training purposes based on fictitious data

Internal Consistency

Description of the Example

Internal consistency reliability was calculated with coefficient alpha, a commonly accepted method of calculating this type of reliability. Students took the test to be placed into an ESL course at their enrolled community college. The alpha calculated for the test was .77 (N = 247). The breakdown of the sample by student race/ethnicity is provided in the table below. The gender makeup of the sample was 57% female and 43% male.

Student Race or Ethnicity	Percent of Sample
White (non-Hispanic)	9%
Hispanic	61%
Black or African American	6%
Asian or Pacific Islander	16%
Native American	8%

Inter-Scorer Reliability (Writing)

<div style="border:1px solid">

<u>Description of the Example</u>

The test battery includes an optional writing test, scored on a five-point scale. In order to be approved to administer the optional writing test, at least four staff members at the school must complete standardized scoring training and demonstrate mastery of the scoring rubric through scoring 20 writing samples of tests taken by students for which the correct score is known. Once staff members within the school have passed the scoring training, the school can begin to administer the optional writing test.

Most student writing tests are scored by only one scorer. However, 15% of the tests are re-read by a second scorer to monitor consistency in scoring. This occurs at all schools administering the writing test, regardless of how long the school has administered the test.

Across all tests administered in the prior academic school year, the consistency between the responses that were second scored was 92% exact or adjacent agreement.

</div>